Sustainable Seas Expeditions

Preliminary Considerations for Data and Information Management

Overview

The following describes several components of data and information management to be considered to support the Sustainable Seas Expeditions (SSE) project. Potential data requirements to plan for and conduct each mission, as well as information products that may be generated during dive operations are also outlined. The primary challenge is to clearly determine the best way in which to manage and distribute information in order to achieve SSE goals. At the very least, this involves:

1. clearly identifying the roles and responsibilities of the project partners to compile, organize, and analyze detailed data using their existing database management systems; and
2. managing specific input and output products.

Ultimately, a data and information management plan will be developed that identifies the tasks required to process, edit, and store information, and which partners will be responsible for ensuring the tasks are completed. The plan will also specify how data, information, and products developed during the expeditions can be accessed.

Data and information management is not about building a database, it is about designing a process and structure for managing information. Providing links to accessible sets of detailed information that are organized and controlled by project partners is critical. This could be accomplished via the SSE web site, which could be structured in a similar manner to some of the NOAA web sites that provide information on projects conducted on board NOAA ships (<http://rho.pmel.noaa.gov/foci/freeman.html> provides one example of how information on projects conducted by the Miller Freeman are disseminated).

Current Protocols - During the expeditions, the Mission Chief will be responsible for the data quality, disposition and archiving of data and samples collected aboard the ship for the primary projects. Copies of these data will be provided on request to participants on the cruise and to any other requesters.

The Commanding Officer will provide the Mission Chief with a single copy of all data collected by ship's personnel. The Mission Chief will provide the Commanding Officer a list of all data collected by the scientific party.

For NOAA sponsored cruises, oceanographic data sets and related information collected are considered to be in the public domain per requirements of the Federal Ocean Data Policy and NOAA Administrative Order 216-101. Principal investigators will have the opportunity to process, analyze or publish their own data sets but it will be understood that data bases will eventually be available to the larger scientific community within 2 years and archived at the National Ocean Data Center (NODC).

Principal Investigators, for each project are required to submit an inventory of data collected within 60 days of collection and a data set for archival within 2 years of collection. Video and still imagery collected during the expeditions shall remain the property of the SSE, but copies will be made available on request to Principal Investigators.

Data Considerations - (Discussed at the Initial Data Management Meeting, 8/19/98)

- Required Data and Information - Planning for and conducting an expedition at any given Sanctuary requires the collection of data and information such as bathymetry, benthic habitats, cultural resources, remotely sensed imagery, and collaborative research and education projects that are relevant to expedition needs.
- Generated Data and Information - Each expedition will generate a wide variety of data and information such as video footage, still photography, quantitative and qualitative observations (physical, chemical, natural resource, etc.) that must be managed effectively in order to make data and information generally available, and to develop information products in a timely and efficient manner.
- Standards and Protocols - Information that is prepared for or generated by an expedition must be in an accessible, usable format. Standards and protocols are an important part of data and information management, covering concerns such as access, timeliness, long-term archiving, and security.
- Ownership/Credit - Agreements concerning ownership and credit for the data and information products generated by the SSE must be specified.
- Databases - It is anticipated that the project will utilize a variety of existing databases, and construct new ones, depending on the specific issue being addressed. For example, the project team is already working on a system to manage information on collaborative research projects. In contrast, efforts are underway to develop a partnership with NASA whereby remote sensing data from the Goddard Distributed Active Archive Center (DAAC) would be processed and made available to the SSE. As noted, the SSE web site can and should be the cornerstone for managing and linking information.

  It is also anticipated that the project will need to develop a specific database management system to handle the data and information generated during the expeditions. The system would be designed to manage the dive and science logs, as well as any other information collected during the dives.
- Evolution - The SSE project is scheduled to run for 5 years. Lessons learned during the first year of deployments will be used to prepare for future work. Therefore, data and information needs, as well as the type of information generated will change over time. By focusing on a process for managing information, the resulting Data and information management plan will be flexible, helping to meet changing conditions and taking advantage of evolving technologies.

Data Types - The following tables list the different types of data and information either required or potentially generated by the SSE project. An important next step will be to refine these lists and begin the process of identifying products, uses, and potential partners.

Table 1. *List of data types required for planning and conducting expeditions.*

			Potential Partners	
Data Type	Data Product	Uses	Lead	Support
Positioning (GPS) information				
SSE activities				
Environmental data (weather/water conditions, etc.)				
Bathymetry				
Baseline benthic habitat information				
Baseline LMR information				
Baseline cultural resource information				
Remote sensing imagery				
Equipment inventory & specifications				
Other projects/programs occurring concurrently at expedition sites				
Agreements (formal & informal)				
Permits				
Contacts				
Collaborative research projects				
Collaborative education efforts				

Table 2. *List of data types potentially generated by the expeditions.*

			Potential Partners	
Data Type	Data Product	Uses	Lead	Support
Video (tape/digital of exploration, operations, interviews, etc.)				
Still photography				
Deployment statistics (location, depth)				
Specimens				
Observations (natural history, technical, hydrographic, physical,				
Research specific information				
tasks completed				
Quick look reports				
Expedition log				
Participants				
Cruise report				

Elements of a Data and Information Management Plan - A data and information management plan should provide the SSE project team and partners with a blueprint of how data and information will be obtained, processed, edited, stored, developed into products, and disseminated. A key component of the plan will be the description of the activities or tasks to be performed and their critical characteristics. The following is a list of the fundamental components of such a plan:

- Goals and Objectives - An outline and description of the goals and objectives specific to data and information management.
- Data and Information Types - A description of:
  - Data required to plan for and conduct expeditions
  - Data generated by the expeditions
- Data Management Activities - Methods for obtaining data and preparing required products, methods for processing and preparing generated products, and methods for disseminating information.
  - Responsibilities - Lead, support
  - Resource requirements - Person weeks, cost
  - Funding sources
  - Schedule
- Data and Information Management Systems
  - List of DBMS operated by partners responsible for managing specific data sets
  - Description of any DBMS to be developed by the SSE project

- Process for Disseminating Information
  - Information products to be disseminated
  - Description of process

Potential Partnerships - As mentioned, project members are currently working with representatives from NASA to determine roles and responsibilities for obtaining, processing, and accessing remote sensing imagery to support the expeditions. In addition, NOAA's National Undersea Research Program (NURP) has recently submitted a proposal to the Environmental Science Data and Information Management (ESDIM) program pertaining to the rescue and digitization of NURP underwater video. The proposal includes the processing of video footage shot during the SSE project. If awarded, this effort will provide the project with significant support.

Next Steps - Once a "data management" team has been assembled, work can begin on the development of a process for managing data and information, and drafting a plan as outlined above. Specifically, the team can focus on developing the partnerships necessary to carry out the work, as well as identifying project specific database needs and designing the overall information management system. This team will be responsible for identifying and evaluating potential models for handling information required for supporting the expeditions, as well as the information generated during each dive.